

Permutation tests for nonparametric detection

José L. Sanz-González¹, Francisco Álvarez-Vaquero², José E. González-García³

ABSTRACT

In this paper, the authors provide a methodology to design nonparametric permutation tests and, in particular, nonparametric rank tests for applications in detection. In the first part of the paper, the authors develop the optimization theory of both permutation and rank tests in the Neyman–Pearson sense; in the second part of the paper, they carry out a comparative performance analysis of the permutation and rank tests (detectors) against the parametric ones in radar applications.

First, a brief review of some contributions on nonparametric tests is realized. Then, the optimum permutation and rank tests are derived. Finally, a performance analysis is realized by Monte-Carlo simulations for the corresponding detectors, and the results are shown in curves of detection probability versus signal-to-noise ratio.

1. Introduction

Statistical signal processing has been applied to communications and radar systems for more than 50 years, and its bases can be found in the statistics developed in the first half of the 20th-century. Most applications are referred to parametric statistics, where distributions are well characterized by families of parametric distributions. On the other hand, the theory of nonparametric statistics is rarely applied to signal processing, although it has also been developed before the 1960s. A clear application of nonparametric statistics is in radar detection, where nonparametric families of distributions can adequately characterize the noise and clutter.

First, we point out that Savage [1] compiled bibliographic references about nonparametric statistics, and Kassam [2] compiled references on nonparametric applications to communications and radar detection. Most contributions on nonparametric statistics are referred to rank tests [3–6]. Basic

introductions to rank tests are given in [3,4], and more advanced treatments are in [5,6]. Also, the rank tests can be considered as a particular case of the family of permutation tests [6–8]. Therefore, the optimum permutation test [7] is more powerful than the optimum rank test [6], although the former has much more computational complexity than the latter.

Permutation tests (PT) were proposed first by Fisher in 1935 [9], under the name of randomization test. In 1937, Pitman developed first Fisher's ideas, and applied them to the two-sample case, linear regression and block designs [10]. Other equivalent names used in the statistical literature are Fisher's exact tests, re-randomization tests, conditional tests, and permutation tests. Most contributions on PT until 1950 are summarized in [10–12] and references therein. Lehmann and Stein [11] provide the bases for optimizing the PT in the Neyman–Pearson sense (i.e., the most powerful permutation test under a simple alternative hypothesis). Hoeffding [12] demonstrates that many PT are asymptotically as powerful as their parametric counterparts. For analyzing contributions on nonparametric permutation tests in the 2nd half of the 20th century, we can consult books [3–9], papers [12–15], and references therein. After the 1950s, fundamental contributions to nonparametric

statistics have been scarce. For analyzing recent novelties on PT, it is interesting to consult the recent statistical literature, see for example [13–15], where [13] is referred to robustness of PT, [14] applies concepts of invariance, exchangeability and sufficiency to PT, and [15] is referred to the PT computation or PT implementation, being an important issue for practical applications.

In the statistical literature, the implicit applications of nonparametric tests were on biology, agriculture or social sciences for obvious reasons. Nevertheless, nowadays, nonparametric tests have also potential applications to detection in communications and radar systems, as the underlying distributions cannot be well characterized, and only very general properties of the distributions can be established (e.g., symmetry properties). For example, in radar detection under general clutter conditions, the multi-dimensional distribution of a block of clutter samples can have the property of invariance under the permutation of its arguments; so, using a permutation test (or a rank test) we can obtain a “constant false-alarm rate” (CFAR) characteristic for the corresponding detector. However, only a few books can be found on nonparametric statistics applied to radar or communications [16,17]. Interesting classical tutorial papers about nonparametric detection are due to Carlyle and Thomas [18] and Thomas [19]. Applications of rank tests to radar detection can be found in [20–24] and references therein. Other rank applications to (wireless) communications are, for example, in [25–29] and references therein. Finally, although rank tests have been applied to detection [16–29], the more general family of permutation tests has seldom been applied to radar detection [30–35], at least, as far as the authors know.

Now, some comments on the methodology of test optimization are required. As is well known, when the hypotheses have nuisance parameters, often, the uniformly most powerful (UMP) test does not exist. Then, it is convenient to explore the existence of the UMP unbiased test [7] or, alternatively, to explore the existence of the UMP invariant test (based on an invariance property under a group of transformations) [7,12–14,36–38]. In this paper, our methodology of designing optimum tests is based on distribution-free statistics over a nonparametric family of distributions, which provides automatically the CFAR characteristic of our detectors under very general noise (or clutter) conditions.

The paper is organized as follows. In Section 2, we obtain the optimum permutation test in the Neyman–Pearson sense, following a new methodology. In Section 3, the rank test is introduced as a subfamily of the general permutation test in a novelty approach. Then, in Section 4, we apply the theory developed in Section 2 to radar detection, resulting in the optimum permutation detector. Section 5 includes computer simulation results of permutation and rank detectors against parametric detectors. Finally, Section 6 summarizes the main conclusions of the paper.

2. Optimum permutation tests

As is well known, a parametric family of distributions is defined by a finite number of parameters (e.g., the

family of Gaussian distributions); on the contrary, a nonparametric family of distributions [3–8] cannot be defined by a finite number of parameters (e.g., the family of all continuous distributions).

Suppose that hypothesis H_0 is the nonparametric family of n -dimensional distributions with the property of invariance under the permutation of their n -components (also, known as the “exchangeability property” [4]), i.e., if $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$, then

$$f_{\mathbf{x}}(x_1, x_2, \dots, x_n) = f_{\mathbf{x}}(x_2, x_1, \dots, x_n) = \dots = n! \text{ permutations} \dots = f_{\mathbf{x}}(x_n, x_{n-1}, \dots, x_1), \quad (1)$$

where $f_{\mathbf{x}}(\cdot)$ is the n -dimensional probability density function of the random vector $\mathbf{X} = (X_1, X_2, \dots, X_n)$. Property (1) is also satisfied by the distribution function $F_{\mathbf{x}}(\cdot)$. If the random vector \mathbf{X} has independent and identically distributed components, property (1) is always fulfilled.

A permutation test is defined by means of a statistic $\tau(\mathbf{x})$ as follows. Suppose we have got a sample vector $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$ with $x_i \neq x_j$ for $i \neq j$, $i, j = 1, 2, \dots, n$; then, compute $\tau(\cdot)$ for each permutation \mathbf{x}_p of \mathbf{x} . Define $\tau_1(\mathbf{x})$ as the first greatest (maximum) value of $\tau(\cdot)$ among all permutations of \mathbf{x} ; also, define $\tau_2(\mathbf{x})$ as the second greatest value of $\tau(\cdot)$ among all permutations of \mathbf{x} ; and so on, until the smallest (minimum) value $\tau_{n!}(\mathbf{x})$ over all permutations of \mathbf{x} . That is, we can write

$$\tau_i(\mathbf{x}) = i^{\text{th}}\text{-greatest}\{\tau(\mathbf{x}_p) : \mathbf{x}_p \text{ is a permutation of } \mathbf{x} = (x_1, x_2, \dots, x_n), x_k \neq x_j, k \neq j, k, j = 1, 2, \dots, n\}, \quad i = 1, 2, \dots, n! \quad (2)$$

satisfying

$$\tau_{n!}(\mathbf{x}) < \tau_{n!-1}(\mathbf{x}) < \dots < \tau_k(\mathbf{x}) < \dots < \tau_2(\mathbf{x}) < \tau_1(\mathbf{x}). \quad (3)$$

The statistic $\tau(\mathbf{x})$ is a maximal permutation statistic, if (2) and (3) are satisfied for $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$ with $x_i \neq x_j$ for $i \neq j$, $i, j = 1, 2, \dots, n$, almost everywhere. Note that $\tau(\mathbf{x}) \in \{\tau_1(\mathbf{x}), \tau_2(\mathbf{x}), \dots, \tau_{n!}(\mathbf{x})\}$ and the equation $\tau(\mathbf{x}) = \tau_i(\mathbf{x})$ has a single solution in all permutations of \mathbf{x} if inequalities (3) are fulfilled; i.e., if $\mathbf{x}^{(i)} = \arg(i\text{th-greatest}\{f_{\mathbf{x}}(\mathbf{x}_p | H_1) : \mathbf{x}_p \text{ over all permutations of } \mathbf{x}\})$ and (3) is satisfied, then we have a single $\mathbf{x}^{(i)}$ satisfying the equality: $\tau(\mathbf{x}^{(i)}) = \tau_i(\mathbf{x}^{(i)})$, where $\mathbf{x}^{(i)} \in \{\text{all permutations of } \mathbf{x}\}$.

According to (2) and (3), $\tau(\mathbf{x}) = \tau_i(\mathbf{x})$ defines a region $D_i \subset \mathfrak{R}^n$ for each $i = 1, 2, \dots, n!$, such that

$$D_i = \{\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n : \tau(\mathbf{x}) = \tau_i(\mathbf{x}), x_k \neq x_j, k \neq j, k, j = 1, 2, \dots, n\}, \quad i = 1, 2, \dots, n! \quad (4)$$

The probability measure of regions $D_i \subset \mathfrak{R}^n$, $i = 1, 2, \dots, n!$ is guaranteed by the statistic $\tau(\mathbf{x})$ and the underlying distribution in the \mathfrak{R}^n -space.

Also, we define a region D_0 corresponding to the union of all hyperplanes $x_i = x_j$ for $i \neq j$, $i, j = 1, 2, \dots, n$, as follows $D_0 = \cup \{\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathfrak{R}^n : x_i = x_j, i \neq j, i, j = 1, 2, \dots, n\}$. (5)

The probability measure of region D_0 is zero for continuous distributions in the \mathfrak{R}^n -space.

Definitions (2) and (3) generate a partition on \mathfrak{R}^n . Then, from (4) and (5), we have $D_i \cap D_j = \emptyset$ for $i \neq j$, $i, j = 0, 1, 2, \dots, n!$, and $\cup_{i=0}^{n!} D_i = \mathfrak{R}^n$. Note that if $\mathbf{x} \in D_i$, then $\mathbf{x}_p \notin D_i$

$i=1, 2, \dots, n!$, where \mathbf{x}_p is any permutation of $\mathbf{x}=(x_1, x_2, \dots, x_n) \in \mathcal{R}^n$ with $x_k \neq x_j$ for $k \neq j$, $k, j=1, 2, \dots, n$. Also, for distributions fulfilling property (1), we have $\Pr\{D_i\}=(1/n!)$, $i=1, 2, \dots, n!$, and $\Pr\{D_0\}=0$, because the probability of tied samples ($x_i=x_j$, $i \neq j$) is zero for continuous distributions.

Now, in order to test the null hypothesis H_0 against the alternative hypothesis H_1 , we can use the permutation test, defined as follows

$$\begin{array}{c} H_1 \\ \tau(\mathbf{x}) > \tau_k(\mathbf{x}), \\ H_0 \end{array} \quad (6)$$

where threshold $\tau_k(\mathbf{x})$ is obtained from (2) and (3) after computing the statistic $\tau(\mathbf{x})$ for all permutations of \mathbf{x} , and ordering (sorting) them, from the first greatest value (the maximum value) and denoted as $\tau_1(\mathbf{x})$, following the second greatest value $\tau_2(\mathbf{x})$, and so on, until the k th-greatest value $\tau_k(\mathbf{x})$. The index k is an integer in $1 \leq k \leq n!$ and fixed beforehand, as it is indicated below.

It is worth noting that permutation test (6) can be interpreted as a test with adaptive threshold $\tau_k(\mathbf{x})$, because we have to compute the threshold for each new sample vector \mathbf{x} before comparing it with the statistic $\tau(\mathbf{x})$. In order to obtain threshold $\tau_k(\mathbf{x})$ in (6), we have to compute statistic $\tau(\mathbf{x})$ for all permutations of \mathbf{x} and, after, ordering (sorting) them according to (3). Then, the computation problem of the permutation test is apparent for large n -values (e.g., if $n=16$, we have approximately 2×10^{13} computations of the statistic). In the applications, we will consider this computational problem, and propose the corresponding algorithms.

Suppose hypothesis H_0 (target absent) satisfies property (1) and $f_{\mathbf{x}}(\mathbf{x}|H_0)$ is a probability density function (pdf) in H_0 . The false-alarm probability (P_{fa}) of the test (6) is given by

$$\begin{aligned} P_{fa} &= \Pr\{\tau(\mathbf{X}) \geq \tau_k(\mathbf{X})|H_0\} \\ &= \int_{\mathcal{R}^n} u(\tau(\mathbf{x}) - \tau_k(\mathbf{x})) \cdot f_{\mathbf{x}}(\mathbf{x}|H_0) d\mathbf{x} = \frac{k}{n!}, \end{aligned} \quad (7)$$

where $u(\cdot)$ is the unit-step function, i.e., $u(x)=1$ if $x \geq 0$, and $u(x)=0$ if $x < 0$. Note that (6) is strictly CFAR, because $P_{fa}=k/n!$ for any $f_{\mathbf{x}}(\mathbf{x}|H_0)$ satisfying property (1). The proof of Eq. (7) is very simple, noting that each permutation of $\mathbf{x}=(x_1, x_2, \dots, x_n)$ is equally likely under H_0 , so $\Pr\{\tau(\mathbf{x}) = \tau_i(\mathbf{x})\} = 1/n!$, $i=1, 2, \dots, n!$ and, consequently, $\Pr\{\tau(\mathbf{x}) \geq \tau_k(\mathbf{x})\} = k/n!$ that corresponds to the last equality in Eq. (7). Finally, the index k of $\tau_k(\mathbf{x})$ is obtained from (7) and the required P_{fa} .

Under hypothesis H_1 (target present), the detection probability (P_d) of the test (6) is given by

$$\begin{aligned} P_d &= \Pr\{\tau(\mathbf{X}) \geq \tau_k(\mathbf{X})|H_1\} = \int_{\mathcal{R}^n} u[\tau(\mathbf{x}) - \tau_k(\mathbf{x})] \cdot f_{\mathbf{x}}(\mathbf{x}|H_1) d\mathbf{x} \\ &= E\{u[\tau(\mathbf{X}) - \tau_k(\mathbf{X})]|H_1\}, \end{aligned} \quad (8)$$

where $E\{\cdot\}$ means mathematical expectation and $f_{\mathbf{x}}(\mathbf{x}|H_1)$ is a pdf in H_1 . Also, we are interested in finding statistic $\tau(\mathbf{x})$ that maximizes (8) under constraint (7), i.e.,

permutation test (6) is optimum in the Neyman–Pearson sense. This optimization is provided below in the Main Theorem.

The Main Theorem:

Suppose the null hypothesis H_0 is nonparametric, fulfilling property (1), with $f_{\mathbf{x}}(\mathbf{x}|H_0)$ any pdf in H_0 , and H_1 is a simple alternative hypothesis with a unique pdf $f_{\mathbf{x}}(\mathbf{x}|H_1)$; if $\tau(\mathbf{x})$ is the likelihood ratio statistic, i.e., $\tau(\mathbf{x})=(f_{\mathbf{x}}(\mathbf{x}|H_1)/f_{\mathbf{x}}(\mathbf{x}|H_0))$, then permutation test (6) is optimum in the Neyman–Pearson sense.

Proof. See Appendix I.

An important conclusion of the Main Theorem is that the optimum statistic of the Neyman–Pearson test is also an optimum statistic for the permutation test, under the same conditions of the hypotheses H_0 and H_1 . Nevertheless, the converse is not true in general (see Appendix I).

Definition. We say that two permutation statistics $\tau(\mathbf{x})$ and $\lambda(\mathbf{x})$ are equivalent if they provide the same permutation partition (3) of the \mathcal{R}^n -space. Consequently, the corresponding permutation tests are also equivalent, in the sense of providing the same P_d for a given P_{fa} .

The two following propositions will be useful in a next section on radar applications.

Proposition 1. Suppose that $\tau(\mathbf{x})$ and $\lambda(\mathbf{x})$ are two statistics and $\mathbf{x}_p, \mathbf{x}_q \in \{\text{permutations of } \mathbf{x}\}$, $\mathbf{x} \in \mathcal{R}^n$. Now, for every $\mathbf{x}_p \neq \mathbf{x}_q$ and $\forall \mathbf{x} \in \mathcal{R}^n$, if $\tau(\mathbf{x}_p) < \tau(\mathbf{x}_q) \Leftrightarrow \lambda(\mathbf{x}_p) < \lambda(\mathbf{x}_q)$, then permutation statistics $\tau(\mathbf{x})$ and $\lambda(\mathbf{x})$ are equivalent.

Proof. Applying the condition: $\tau(\mathbf{x}_p) < \tau(\mathbf{x}_q) \Rightarrow \lambda(\mathbf{x}_p) < \lambda(\mathbf{x}_q)$ to the partition defined by (3), we obtain the same partition defined by

$$\lambda_n(\mathbf{x}) < \lambda_{n-1}(\mathbf{x}) < \dots < \lambda_k(\mathbf{x}) < \dots < \lambda_2(\mathbf{x}) < \lambda_1(\mathbf{x}), \quad (9)$$

where

$$\begin{aligned} \lambda_i(\mathbf{x}) &= \text{ith-greatest } \{\lambda(\mathbf{x}_p) : \mathbf{x}_p \text{ over all } \mathbf{x}\text{-permutations}\} \\ i &= 1, 2, \dots, n! \end{aligned} \quad (10)$$

and vice versa. Then, partitions (9) and (3) are identical. Consequently, the permutation tests provided by $\lambda(\mathbf{x})$ and $\tau(\mathbf{x})$ are equivalent. \square

Remark. A sufficient condition for the permutation equivalence of $\lambda(\mathbf{x})$ and $\tau(\mathbf{x})$ is that $\lambda(\mathbf{x})=\phi(\tau(\mathbf{x}))$, being $\phi(\tau)$ a monotonically (strictly) increasing function of τ . This is straightforward from Proposition 1 due to the following relations: $\tau(\mathbf{x}_p) < \tau(\mathbf{x}_q) \Leftrightarrow (\tau(\mathbf{x}_p)) < \phi(\tau(\mathbf{x}_q)) \Leftrightarrow \lambda(\mathbf{x}_p) < \lambda(\mathbf{x}_q)$.

Now, under H_0 , suppose that $f_{\mathbf{x}}(\mathbf{x}|y, H_0)$ is the conditional pdf satisfying property (1) with y a nuisance parameter, and $f_y(y)$ is the pdf of the random parameter Y . Define $f_{\mathbf{x}}(\mathbf{x}|H_0)$ as follows

$$f_{\mathbf{x}}(\mathbf{x}|H_0) = \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}|y, H_0) f_y(y) dy. \quad (11)$$

It is apparent that $f_{\mathbf{x}}(\mathbf{x}|H_0)$ also satisfies property (1).

In a similar way, under H_1 , define

$$f_{\mathbf{x}}(\mathbf{x}|H_1) = \int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}|y, H_1) f_y(y) dy$$

$$= \int_{-\infty}^{\infty} \lambda(\mathbf{x}; y) \cdot f_{\mathbf{x}}(\mathbf{x}|y, H_0) f_Y(y) dy, \quad (12)$$

where

$$\lambda(\mathbf{x}; y) = \frac{f_{\mathbf{x}}(\mathbf{x}|y, H_1)}{f_{\mathbf{x}}(\mathbf{x}|y, H_0)}. \quad (13)$$

Then, from (12) and (11), the likelihood ratio $\lambda(\mathbf{x})$ can be expressed as follows

$$\lambda(\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}|H_1)}{f_{\mathbf{x}}(\mathbf{x}|H_0)} = \frac{\int_{-\infty}^{\infty} \lambda(\mathbf{x}; y) \cdot f_{\mathbf{x}}(\mathbf{x}|y, H_0) f_Y(y) dy}{\int_{-\infty}^{\infty} f_{\mathbf{x}}(\mathbf{x}|y, H_0) f_Y(y) dy}. \quad (14)$$

Proposition 2. Suppose that $\lambda(\mathbf{x}; y) = \phi(\tau(\mathbf{x}); y)$, being $\phi(\tau; y)$ a monotonically increasing function of τ (and y is a fixed parameter), then permutation statistic $\lambda(\mathbf{x}; y)$ is equivalent to $\tau(\mathbf{x})$, and equivalent to permutation statistic $\lambda(\mathbf{x})$, defined in (14).

Proof. The permutation equivalence between $\tau(\mathbf{x})$ and $\lambda(\mathbf{x}; y)$ is a direct consequence of Proposition 1. The permutation equivalence between $\tau(\mathbf{x})$ and $\lambda(\mathbf{x})$ defined in (14) is obtained from Proposition 1 and relation (14) with the fact that both $f_{\mathbf{x}}(\mathbf{x}|H_0)$ and $f_{\mathbf{x}}(\mathbf{x}|y, H_0)$ satisfy property (1). \square

Finally, the Main Theorem and Propositions 1 and 2 are new, at least, as far as the authors know, and they will be used later in the applications.

3. Relationship between permutation tests and rank tests

Partition (4) defines a general permutation partition, in accordance with the general statistic $\tau(\mathbf{x})$. Suppose that, as a particular case, $\tau(\mathbf{x}) = \tau_i$ is constant for any $\mathbf{x} = (x_1, x_2, \dots, x_n)$ satisfying $x_{i_1} < x_{i_2} < \dots < x_{i_n}$ for $i_j \neq i_k$, if $j \neq k$, $j, k = 1, 2, \dots, n$. Furthermore, suppose that τ_i are $n!$ different constants for $i = 1, 2, \dots, n!$. In these conditions, partition (4) can be written as follows

$$D_i = \{\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathcal{R}^n : x_{i_1} < x_{i_2} < \dots < x_{i_n}, i_j \neq i_k, \text{ if } j \neq k, j, k = 1, 2, \dots, n\}, i = 1, 2, \dots, n!. \quad (15)$$

It is supposed that tied samples (i.e., $x_i = x_j$, $i \neq j$) are solved by uniform randomization [23]. Partition (15) is well defined by the rank vector $\mathbf{r} = (r_1, r_2, \dots, r_n)$ where \mathbf{r} is a permutation of $(1, 2, \dots, n)$, and can be computed from sample vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$ after ordering (sorting) its components, i.e., $x_{i_1} < x_{i_2} < \dots < x_{i_n}$, so we have $r_{i_k} = k$, $k = 1, 2, \dots, n$. Note that each value of the rank vector \mathbf{r} defines a unique region D_i , $i = 1, 2, \dots, n!$, shown in (15); for example, $\mathbf{r} = (1, 2, \dots, n)$ corresponds to the region $x_1 < x_2 < \dots < x_n$, and $\mathbf{r} = (2, 1, \dots, n)$ corresponds to $x_2 < x_1 < \dots < x_n$, etc.

Other way of computing ranks is as follows [22,23]

$$\mathbf{r} = (r_1, r_2, \dots, r_n); r_i = \sum_{k=1}^n u(x_i - x_k), \quad i = 1, 2, \dots, n$$

$$1 \leq r_i \leq n. \quad (16)$$

Notice that r_i is the rank of sample x_i , $i = 1, 2, \dots, n$, corresponding to the order number of x_i in the block of

the n components of the sample vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$. On the other hand, $\mathbf{r} = (r_1, r_2, \dots, r_n)$ is the rank vector of the sample vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$, where the components of the rank vector are the ranks of the samples.

If property (1) is fulfilled, each region D_i of partition (15) has probability measure equals to $1/n!$ for $i = 1, 2, \dots, n!$; consequently, the rank vector $\mathbf{r} = (r_1, r_2, \dots, r_n)$ is uniformly distributed in its $n!$ points. Furthermore, the rank vector is a distribution-free statistic under the nonparametric family of distributions defined by (1), i.e., the distribution of the rank vector is independent of the underlying distribution, whenever property (1) is satisfied.

The partition given by (15) is equivalent to the rank vector (16). So, the optimum rank test, in the Neyman-Pearson sense, can be obtained from the rank vector probabilities as follows.

$$\Pr(\mathbf{R} = \mathbf{r} | H_0) = \frac{1}{n!}, \quad (17)$$

$$\Pr(\mathbf{R} = \mathbf{r} | H_1) = \Pr(X_{i_1} < X_{i_2} < \dots < X_{i_n} | H_1) = \int_{x_{i_1} < x_{i_2} < \dots < x_{i_n}} f_{\mathbf{x}}(\mathbf{x} | H_1) d\mathbf{x}, \quad (18)$$

where $\mathbf{r} = (r_1, r_2, \dots, r_n) \in \{\text{all permutations of } (1, 2, \dots, n)\}$. Consequently, the optimum rank test is derived from the Neyman-Pearson lemma [6,7,39,40]

$$\lambda(\mathbf{r}) = \frac{\Pr(\mathbf{R} = \mathbf{r} | H_1)}{\Pr(\mathbf{R} = \mathbf{r} | H_0)} = n! \cdot \Pr(\mathbf{R} = \mathbf{r} | H_1) \underset{H_0}{\overset{H_1}{>}} \lambda_0, \quad (19)$$

where λ_0 is the threshold (a fixed value), and $\lambda(\mathbf{r})$ is the rank test statistic that, in general, it does not have a closed-form expression, and a computable suboptimum rank statistic is more practical (e.g., the well-known Wilcoxon rank statistic [3–6]).

Finally, according to the above, the family of rank test can be considered as a subfamily of the family of permutation tests. Hence, the optimum permutation test is more powerful than the optimum rank test; but, permutation tests are much more computational complex than rank tests.

4. Permutation tests in radar detection

Now, the permutation test theory developed in Section 2 will be applied to radar detection. We suppose that the signal comes from a two-dimensional pulsed-radar system. In order to test hypothesis H_0 (target absent) against hypothesis H_1 (target present) for each azimuth in a specific range cell, we take M reference samples corresponding to the range cells surrounding the cell under test. Also, we consider a block of N pulses for each azimuth (or the block of pulses per antenna beamwidth in a radar system with a mechanically steerable antenna); then, for each pulse i , we have the row vector of samples $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$, $i = 1, 2, \dots, N$, (20)

where the last component x_i of vector \mathbf{x}_i is the sample of the range cell under test. Because we consider a

non-coherent detection approach, we suppose that our data are samples of the linear envelope (the phase is discarded). In order to simplify the notation, let us define the probability density functions (pdfs) $f_{0i}(x) = f_{x_i}(x|H_0)$ and $f_{1i}(x) = f_{x_i}(x|H_1)$. Also, $F_{0i}(x)$ and $F_{1i}(x)$ are the corresponding distribution functions.

4.1. IID clutter model

In a first analysis, the two hypotheses H_0 (target absent) and H_1 (target present) are defined in terms of probability density functions, which are as follows for nonfluctuating or Swerling II target models [22,24]

$$H_0 : f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_0) = \prod_{i=1}^N f_{0i}(x_i) \prod_{j=1}^M f_{0i}(x_{ij}) \quad (21)$$

$$H_1 : f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_1) = \prod_{i=1}^N f_{1i}(x_i) \prod_{j=1}^M f_{0i}(x_{ij}) \quad (22)$$

Note that under H_0 , the components of the random vector $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$ are independent and identically distributed (IID) for each pulse i ($i=1, 2, \dots, N$); under H_1 , only the reference samples $x_{i1}, x_{i2}, \dots, x_{iM}$ are IID. Also, vectors \mathbf{X}_i , $i=1, 2, \dots, N$, are mutually independent. Hypothesis H_0 of (21) is IID by blocks, i.e., it is IID inside each random vector $\mathbf{X}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$, but random vectors \mathbf{X}_i , $i=1, 2, \dots, N$, are not IID among them. So, $f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_0)$ satisfies property (1) in each vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$, $i=1, 2, \dots, N$.

According to the Main Theorem of Section 2, the optimum permutation test in the Neyman–Pearson sense is obtained from the likelihood ratio of (22) and (21): $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$. Now, if we apply permutations in each vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$, $i=1, 2, \dots, N$, then we obtain $[(M+1)]^N$ different values of the statistic $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$, and compare it with the number $[(M+1)]^N$ that corresponds to all blocks of permutations. Also, we define $\lambda_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ for computing the adaptive threshold, as follows

$$\lambda_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{\prod_{i=1}^N f_{1i}(x_{ik_i}) \prod_{j=1}^{M+1} f_{0i}(x_{ij})}{\prod_{i=1}^N f_{0i}(x_{ik_i}) \prod_{j=1}^{M+1} f_{0i}(x_{ij})} = \prod_{i=1}^N \frac{f_{1i}(x_{ik_i})}{f_{0i}(x_{ik_i})}, \quad (23)$$

where $\mathbf{k} = (k_1, k_2, \dots, k_N)$, being $k_i = 1, 2, \dots, M+1$, and $i=1, 2, \dots, N$. Notice that the last component of (20) is $x_i = x_{i, M+1}$, $i=1, 2, \dots, N$. Also, $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is included in (23) as a particular case, i.e., $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \lambda_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ for $\mathbf{k} = (M+1, M+1, \dots, M+1)$.

On the other hand, the application of the Neyman–Pearson lemma to hypotheses (21) and (22), considering the log-likelihood ratio test, leads us to the following optimum parametric test

$$T(\mathbf{x}) = \sum_{i=1}^N a_i(x_i) \begin{matrix} > \\ < \end{matrix} T_0; \quad a_i(x_i) = \log \left(\frac{f_{1i}(x_i)}{f_{0i}(x_i)} \right), \quad (24)$$

where T_0 is the threshold of the test, $f_{1i}(x)$ and $f_{0i}(x)$ are the pdfs of the sample x_i under H_1 and H_0 , respectively.

Column vector $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$ is composed of the N samples under test, and statistic $T(\mathbf{x}) = \log(\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N))$.

Also, for convenience, (23) is transformed into the equivalent form

$$T_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \log(\lambda_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)) = \log \left(\prod_{i=1}^N \frac{f_{1i}(x_{ik_i})}{f_{0i}(x_{ik_i})} \right) = \sum_{i=1}^N a_i(x_{ik_i}), \quad (25)$$

where $a_i(x)$ is defined in (24), and $T(\mathbf{x})$ is included in (25) for $\mathbf{k} = (M+1, M+1, \dots, M+1)$.

For a block of N pulses, we have $(M+1)^N$ possible values of $T_{\mathbf{k}}(\cdot)$ given by (25), which are equally likely under hypothesis H_0 defined in (21). Now, we order (sort) $T_{\mathbf{k}}(\cdot)$ from its smallest value to its greatest value, as follows

$$\inf_{\mathbf{k}} (T_{\mathbf{k}}(\cdot)) = (T_{\mathbf{k}}(\cdot))_{(M+1)^N} < (T_{\mathbf{k}}(\cdot))_{(M+1)^N-1} < \dots < (T_{\mathbf{k}}(\cdot))_K < \dots < (T_{\mathbf{k}}(\cdot))_1 = \sup_{\mathbf{k}} (T_{\mathbf{k}}(\cdot)). \quad (26)$$

Finally, consider the threshold T_0 in (24) as given by $T_0 = (T_{\mathbf{k}}(\cdot))_K$, (27)

where K is the number of $T_{\mathbf{k}}(\cdot)$ -values which are greater than or equal to T_0 , and it is easily obtained from (26) for low K -values. So the permutation test is realized by

$$\begin{matrix} H_1 \\ T(\mathbf{x}) > \\ < \end{matrix} (T_{\mathbf{k}}(\cdot))_K, \quad H_0 \quad (28)$$

where $T(\mathbf{x})$ is the test statistic of the samples under test, given by (24).

In [34], we have described two permutation test algorithms for computing (28). These algorithms can be implemented in workstations or in personal computers. The first algorithm (ALGORITHM_1) is based on direct computations of (25) to establish (26). The second algorithm (ALGORITHM_2) is based on the computation of $T_{\mathbf{k}}(\cdot)$ -histogram of (25), by realizing convolutions of N histograms corresponding to N rows of data (see [34] for the details).

The P_{fa} of the permutation test (28) is given by

$$P_{fa} = \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | H_0\} = \frac{K}{(M+1)^N}. \quad (29)$$

We point out that K is computed from the nominal P_{fa} , i.e., $K = \lceil P_{fa} \cdot (M+1)^N \rceil$.

The P_d of the permutation test (28) is given by

$$P_d = \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | H_1\} = \int u[T(\mathbf{x}) - (T_{\mathbf{k}}(\cdot))_K] f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_1) d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N, \quad (30)$$

where $u[\cdot]$ is the unit-step function, and $f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_1)$ is given by (22).

4.2. SIRP clutter model

The spherically invariant random process (SIRP), as clutter model for the range cell under test and a block of N pulses

[43], can be represented by the compound model: $\tilde{\mathbf{X}} = \tilde{\mathbf{W}} \cdot Y$, where $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{W}}$ are N -dimensional complex random column vectors, and Y is a real random scalar (independent of $\tilde{\mathbf{W}}$). Also, $\tilde{\mathbf{W}} = (\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_N)^T$ is a complex-Gaussian random column vector, and the components $\tilde{W}_i, i=1, 2, \dots, N$, are mutually independent if the radar system uses frequency agility from pulse-to-pulse. Without loss of generality, we can suppose that each component of \tilde{W}_i has unit power (or \tilde{W}_i has power 2). Moreover, if the detector uses M range reference cells (surrounding the cell under test), we shall have the column vectors $\tilde{\mathbf{X}}^{(j)} = \tilde{\mathbf{W}}^{(j)} \cdot Y^{(j)}, j=1, 2, \dots, M$; also, an important particular case is $Y^{(j)}=Y, j=1, 2, \dots, M$, corresponding to a case of (locally) homogenous clutter in the reference cells. In this paper, we are referring to the non-coherent detection; then, the linear envelope $X = |\tilde{X}|$, where \tilde{X} is the complex envelope, and $X = W \cdot Y$, where W is a Rayleigh random variable and Y is a positive real random scalar (independent of W) with a very general probability density function. In Appendix II, it is proved that the optimum (or suboptimum) permutation test statistic under Gaussian noise (linear or quadratic statistic) is also optimum (or suboptimum) under SIRP clutter. Hence, the permutation test (28) for the IID case is applied in the SIRP case with the corresponding changes in the test statistics $T(\mathbf{x})$ and $T_{\mathbf{k}}(\cdot)$.

5. Detector performance analysis

In this section, we present a sample of the results of P_d versus signal-to-clutter ratio (SCR) for some optimum non-parametric permutation (and rank [22,24]) tests (detectors) against optimum parametric detectors [32] under K-distributed clutter and nonfluctuating (NF) and Swerling II (SWII) target models. Additional results can be found in [42,43]. The clutter shape parameter (ν) has been $\nu=0.5$ as a typical value for spiky clutter, and $\nu=\infty$ for Rayleigh clutter. Also, we have considered two clutter models: an ideal case of independent and identically distributed clutter samples, and another more realistic case of spherically invariant random process clutter model with the exchangeability property. The performance analysis was realized by Monte Carlo simulations on a personal computer.

Finally, some words are required about the computational complexity of the detectors. Parametric detectors (24) have a little computational complexity: the number of operations (additions and multiplications) is of the order of the number of samples (N). On the other hand, rank detectors [22,24] applied to the data (20) have a computational complexity proportional to $(M+1) \cdot N$. However, the permutation test [34] realized by a direct algorithm (ALGORITHM_1) has a computational complexity proportional to $(M+1)^N$, and realized by an algorithm based on histograms (ALGORITHM_2) has a computational complexity proportional to $(M+1)^2 \cdot N$ or, even less, if an FFT-algorithm is used: $N \cdot (M+1) \cdot \log_2(M+1)$. For more details, see [34].

5.1. Analysis under IID (K-distributed) clutter samples

According to our previous works on parametric and nonparametric (rank and permutation) detectors [22–24,32] under IID clutter samples, the optimum statistic for parametric and permutation detectors is the clipping statistic [32]

under spiky K-distributed clutter ($\nu=0.5$) and the linear (or quadratic) statistic under Rayleigh clutter ($\nu=\infty$). For rank detectors [22,24], the linear rank statistic is quasi-optimum under spiky clutter ($\nu=0.5$) and suboptimum under Rayleigh clutter ($\nu=\infty$).

In Fig. 1, we show a sample of the results in terms of P_d versus SCR for optimum parametric, permutation and rank detectors with some different detector parameters, under Rayleigh clutter (shape parameter: $\nu=\infty$), and the NF target model. The best detector is the parametric one, followed by the permutation detector and the rank one, respectively. If we consider $P_d=0.8$, permutation detector is about 1 dB away from parametric one for $N=8$ (i.e., the loss $L \approx 1$ dB); and $L \approx 0.5$ dB for $N=16$ or 32. Also, if $P_d=0.8$, the rank has a loss $L \approx 2$ dB for $N=8$, and $L \approx 1$ dB for $N=16$ or 32. Similar tendencies are observed for the SWII case when it is compared with the NF case, although the SCR required in the SWII case is greater than the SCR required in the NF case for the same P_d . Other results have been published in [42,43].

From results not shown here (see [32,34,42,43]), for spiky clutter ($\nu=0.5$), the degradation in SCR of each detector when P_{fa} changes from 10^{-3} to 10^{-6} is about 5 dB, and from $P_{fa}=10^{-6}$ to 10^{-8} , the degradation is about 2 dB. On the other hand, for Rayleigh clutter, there are SCR-degradations of 2 dB approximately from $P_{fa}=10^{-3}$ to 10^{-6} , and 1 dB approximately from $P_{fa}=10^{-6}$ to 10^{-8} . Finally, the permutation detectors are low sensitive with respect to M -variations, e.g., a gain less than 0.7 dB is attained from $M=7$ to $M=15$ with the other parameter fixed. Note that as $N \rightarrow \infty$ and $M \rightarrow \infty$, the optimum permutation and rank detectors attain the same P_d of the optimum parametric detector.

5.2. Analysis under SIRP (K-distributed) clutter

Now, we analyze detector performances under SIRP clutter with K-distribution. According to the SIRP case considered in Section 4.2, the optimum permutation test

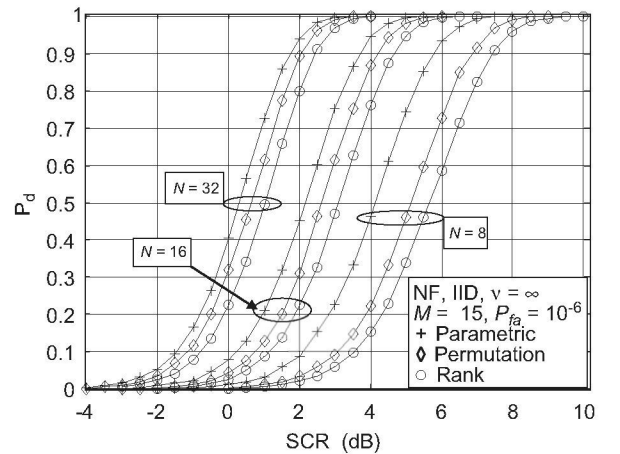


Fig. 1. Detection probability (P_d) versus Signal-to-Clutter Ratio (SCR) for optimum detectors: parametric (+), permutation (◇) and rank (○), under nonfluctuating target model (NF) and IID Rayleigh clutter ($\nu=\infty$). Parameters: $M=15$; $N=8, 16$ and 32 ; $P_{fa}=10^{-6}$.

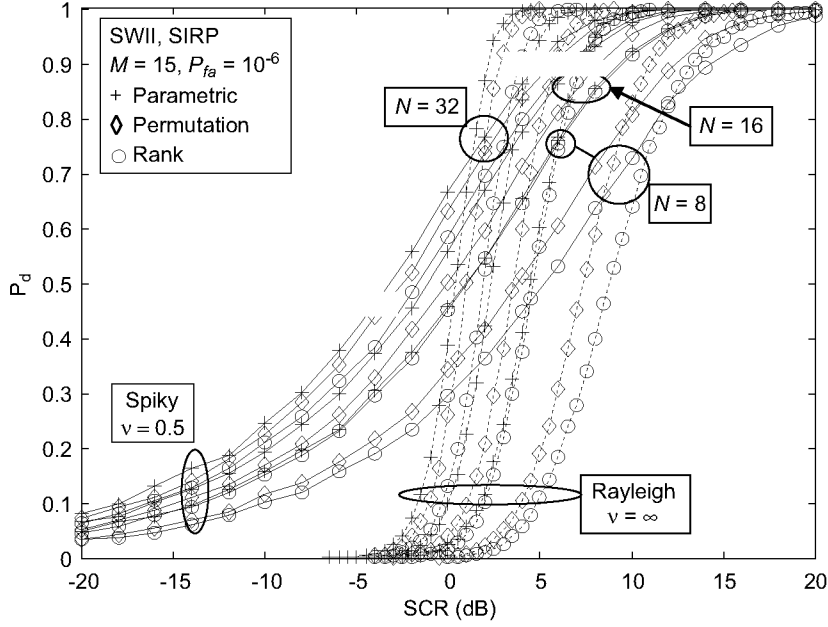


Fig. 2. Detection probability (P_d) versus Signal-to-Clutter Ratio (SCR) for optimum detectors: parametric (+), permutation (◇) and rank (○), under Swerling II target model (SWII), and SIRP K-distributed spiky clutter ($v=0.5$, continuous lines) and Rayleigh clutter ($v=\infty$, dotted lines). Parameters: $M=15$; $N=8, 16$ and 32 ; $P_{fa}=10^{-6}$.

statistics correspond to the Gaussian IID case, i.e., linear under NF target models and quadratic under SWII target models.

Fig. 2 shows a sample of the results in P_d vs. SCR curves for optimum detectors: parametric, permutation and rank, under spiky and Rayleigh clutter, all in Fig. 2 for easy comparisons. Other results have been published in [34,43]. We use the following symbols: continuous lines for denoting SIRP (K-distributed) clutter with $v=0.5$, discontinuous (dotted) lines for denoting Rayleigh clutter; lines with “+” for the optimum parametric detector, lines with “◇” for the optimum permutation detector, and lines with “o” for the optimum rank detector. As expected, under the same conditions, the best detector is the parametric one, followed by the permutation detector and, then, the rank one. Under Rayleigh clutter, P_d -curves of all detectors increase rapidly from $P_d=0.1$ to $P_d=0.95$ as SCR increases (moreover, $P_d \approx P_{fa}$ as $SCR < -20$ dB); on the other hand, under SIRP (K-distributed) spiky clutter, P_d -curves of all detectors increase slowly as SCR increases (now, $P_d \approx P_{fa}$ as $SCR < 20 \log_{10}(P_{fa})$ dB). Also, it can be observed that the separation between two curves under spiky clutter is approximately similar to the separation between the two corresponding curves under Rayleigh clutter.

More precisely, Fig. 2 shows curves of optimum detectors with the parameters $M=15$ and $P_{fa}=10^{-6}$, and three N -values (8, 16 and 32) for SWII target model. Note that the detectability gain of any detector from $N=16$ to $N=32$ is approximately 2 dB (less than the 3 dB of a coherent detector). Also, for $N=32$, permutation and rank detectors are very close to parametric detector for NF target model (losses $L < 0.5$ dB). Finally, remember that as

$N \rightarrow \infty$ and $M \rightarrow \infty$, the optimum permutation and rank detectors attain the same P_d of the optimum parametric detector (in fact, the asymptotic conditions accomplish approximately for $N \geq 32$ and $M \geq 15$), under the same conditions of the other parameters.

6. Conclusions and further work

Permutation tests perform as CFAR detectors if hypothesis H_0 (target absent) satisfies the exchangeability property; that is, the multidimensional distribution of a block of clutter samples is symmetric about the permutation of its arguments. The main problem of the permutation tests is the high computational complexity of their implementations; nevertheless, due to the advances in signal processing, it is a solvable problem, even, in real-time applications. On the other hand, the rank tests have low computational complexity, but they are less powerful than permutation tests. Parametric tests have very low computational complexity, but they are not robust under hypotheses variations.

In this paper, the optimum permutation test has been established and applied to radar detection. Computer simulation results have been obtained and plotted in curves of P_d vs. SCR under K-distributed clutter and nonfluctuating and Swerling II target models, considering IID and SIRP (with the exchangeability property) clutter models. The IID clutter model can be considered as the ideal case, and it is the benchmark for comparing against other clutter models. Finally, a sample of the results is shown and discussed in the paper. A general conclusion is that the optimum permutation test has a low detectability loss with respect to its optimum parametric counterpart

for a moderate number of radar data, but the former is more robust than the latter when hypotheses depart from the assumptions.

In order to cope with correlated and non-homogeneous clutter, an important issue is the pre-processing of the radar signal (e.g., clutter filtering, frequency agility, etc.) in order to accomplish the exchangeability property, before permuting or ranking the received samples. Further research work is required for extending the CFAR characteristic of the permutation detectors over other nonparametric clutter models, satisfying some invariance property, and being useful in applications.

Acknowledgments

The authors are very grateful to the reviewers for their comments and suggestions that, indeed, have improved the quality of the paper.

This work has been partially supported by the “Dirección General de Investigación” of the “Ministerio de Ciencia e Innovación” under Grants: TEC2009-14219-C03-01 and TEC2010-21217-C02-02/TCM, and the “Universidad Politécnica de Madrid”, Spain.

Appendix I. Proof of the Main Theorem

Suppose that hypothesis H_0 is the nonparametric family of distributions defined by property (1), and hypothesis H_1 is define by only one distribution that it does not satisfy property (1), say $f_{\mathbf{x}}(\mathbf{x}|H_1)$ is the pdf of $\mathbf{x}=(x_1, x_2, \dots, x_n)$ under hypothesis H_1 . The critical region, defined by $\tau(\mathbf{x}) \geq \tau_k(\mathbf{x})$, of the optimum test in the Neyman–Pearson sense must include some regions D_i , $i=1, 2, \dots, n!$, of the partition (4) with the maximum probability measure under hypothesis H_1 . Then, the optimum permutation partition must generate regions D_i , $i=1, 2, \dots, n!$, in (4) as follows: a region (say, D_1) with the maximum possible probability measure under H_1 , a second region (say, D_2) with the second greatest probability measure, a third region (say, D_3) with the third greatest probability measure, and so on. More precisely, the way of generating the optimum permutation partition is as follows.

For every $\mathbf{x}=(x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$ with $x_i \neq x_j$ for $i \neq j$, $i, j=1, 2, \dots, n$, D_1 is the region that includes \mathbf{x}_1 corresponding to the greatest value of $f_{\mathbf{x}}(\mathbf{x}|H_1)$ among all permutations of \mathbf{x} , in such a way that $\mathbf{x}_1 \in D_1$ and $\mathbf{x}_p \notin D_1$, being \mathbf{x}_p any permutation of \mathbf{x}_1 ; in the same way, D_2 is the region of \mathfrak{R}^n corresponding to the second greatest values of $f_{\mathbf{x}}(\mathbf{x}|H_1)$ among all permutations of \mathbf{x} ; and so on, until the last region $D_{n!}$, composed of points $\mathbf{x} \in \mathfrak{R}^n$ with the smallest value of $f_{\mathbf{x}}(\mathbf{x}|H_1)$ among all permutations of \mathbf{x} , so that if $\mathbf{x} \in D_{n!}$ then $\mathbf{x}_p \notin D_{n!}$, being \mathbf{x}_p any \mathbf{x} -permutation.

Consequently, $\Pr(D_1|H_1) > \Pr(D_2|H_1) > \dots > \Pr(D_{n!}|H_1)$ and, if $\tau(\mathbf{x})=f_{\mathbf{x}}(\mathbf{x}|H_1)$ accomplishes (2) and (3) from (4) and (1), we have $\Pr(D_i|H_0) = (1/n!)$, $i=1, 2, \dots, n!$. So if we take the union of the first k regions as the critical region of the test (region favorable to the H_1 -decision), we have

$$P_{fa} = (k/n!), \text{ and}$$

$$P_d = \Pr\left(\bigcup_{i=1}^k D_i \middle| H_1\right) = \sum_{i=1}^k \Pr(D_i|H_1)$$

is its maximum possible value.

The implementation of the optimum permutation test [7,11] is done by means of the statistic $\tau(\mathbf{x})=f_{\mathbf{x}}(\mathbf{x}|H_1)$, as follows. Suppose we have received a sample vector $\mathbf{x}=(x_1, x_2, \dots, x_n) \in \mathfrak{R}^n$ with $x_i \neq x_j$ for $i \neq j$, $i, j=1, 2, \dots, n$, and writing (2) and (3) here for easy reference, we have

$$\begin{aligned} \tau_i(\mathbf{x}) &= \text{ith-greatest } \{\tau(\mathbf{x}_p) : \mathbf{x}_p \text{ over all} \\ &\mathbf{x}=(x_1, x_2, \dots, x_n)\text{-permutations, } x_k \neq x_j, k \neq j, k, j=1, 2, \dots, n\}, \\ &i=1, 2, \dots, n! \end{aligned} \quad (31)$$

and

$$\tau_{n!}(\mathbf{x}) < \tau_{n!-1}(\mathbf{x}) < \dots < \tau_k(\mathbf{x}) < \dots < \tau_2(\mathbf{x}) < \tau_1(\mathbf{x}). \quad (32)$$

A problem arises if tied values occur in (32), i.e., $\tau_i(\mathbf{x})=\tau_j(\mathbf{x})$, $i \neq j$, due to $\tau(\mathbf{x})=f_{\mathbf{x}}(\mathbf{x}|H_1)$ is not a maximal permutation statistic, or due to there are tied components ($x_i=x_j$, $i \neq j$) in the sample vector. This problem can be solved by uniform randomization among tied values [23]. In fact, ties occur with discrete distributions and, consequently, in computer realization of the permutation tests.

Now, as $f_{\mathbf{x}}(\mathbf{x}|H_0)$ is invariant under permutations of its arguments $\mathbf{x}=(x_1, x_2, \dots, x_n)$, inequalities (32) can be written as

$$\lambda_{n!}(\mathbf{x}) < \lambda_{n!-1}(\mathbf{x}) < \dots < \lambda_k(\mathbf{x}) < \dots < \lambda_2(\mathbf{x}) < \lambda_1(\mathbf{x}), \quad (33)$$

where

$$\begin{aligned} \lambda_i(\mathbf{x}) &= \frac{\tau_i(\mathbf{x})}{f_{\mathbf{x}}(\mathbf{x}|H_0)} \\ &= \text{ith-greatest } \left\{ \frac{f_{\mathbf{x}}(\mathbf{x}_p|H_1)}{f_{\mathbf{x}}(\mathbf{x}_p|H_0)} : \mathbf{x}_p \text{ over all } \mathbf{x}\text{-permutations} \right\} \\ &i=1, 2, \dots, n! \end{aligned} \quad (34)$$

and a permutation test is given by

$$\begin{aligned} &H_1 \\ &\lambda(\mathbf{x}) > \lambda_k(\mathbf{x}), \\ &H_0 \end{aligned} \quad (35)$$

where the statistic $\lambda(\mathbf{x})$ is the likelihood ratio

$$\lambda(\mathbf{x}) = \frac{f_{\mathbf{x}}(\mathbf{x}|H_1)}{f_{\mathbf{x}}(\mathbf{x}|H_0)} \quad (36)$$

and $\lambda_k(\mathbf{x})$ is the adaptive threshold, obtained from $\lambda(\mathbf{x})$ by permuting its arguments, computing $\lambda(\cdot)$ and ordering them according to (33).

An important conclusion of this Appendix is that the optimum statistic of the Neyman–Pearson test: $\lambda(\mathbf{x})=(f_{\mathbf{x}}(\mathbf{x}|H_1)/f_{\mathbf{x}}(\mathbf{x}|H_0))$ to test $f_{\mathbf{x}}(\mathbf{x}|H_0)$ versus $f_{\mathbf{x}}(\mathbf{x}|H_1)$, is also the optimum permutation test statistic $\lambda(\mathbf{x})$, given in (35) and (36). However, the converse is not true, because $\tau(\mathbf{x})=f_{\mathbf{x}}(\mathbf{x}|H_1)$ is also an optimum permutation test statistic according to (32), but it is not an optimum statistic for the Neyman–Pearson test. Furthermore, we also conclude that $\tau(\mathbf{x})=f_{\mathbf{x}}(\mathbf{x}|H_1)$ and $\lambda(\mathbf{x})=(f_{\mathbf{x}}(\mathbf{x}|H_1)/f_{\mathbf{x}}(\mathbf{x}|H_0))$ are

equivalent permutation test statistics, because both generate the same permutation partition.

Appendix II. Optimum permutation statistics under SIRP clutter model

Under hypothesis H_0 , if $X_i = W_i \cdot Y$ and $X_{ij} = W_{ij} \cdot Y$, $j=1, 2, \dots, M$ and $i=1, 2, \dots, N$, then we can write $\mathbf{X}_i = \mathbf{W}_i \cdot Y$, being $\mathbf{W}_i = (W_{i1}, W_{i2}, \dots, W_{iM}, W_i)$, $i=1, 2, \dots, N$. Now, assuming Y known, the conditional probability density functions under the respective hypotheses H_0 (target absent) and H_1 (target present) are defined for nonfluctuating and Swerling II target models as follows

$$f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_k) = \prod_{i=1}^N f_{ki}(x_i | y) \prod_{j=1}^M f_{0i}(x_{ij} | y), \quad k=0,1, \quad (37)$$

where $f_{0i}(x|y)$ is the conditional (assuming y fixed) pdf of a clutter reference sample x in pulse i ($i=1, 2, \dots, N$), and $f_{1i}(x|y)$ is the conditional pdf of a sample x , corresponding to target signal return and clutter in the cell under test for pulse i ($i=1, 2, \dots, N$).

Now, if the probability density function of the positive random variable Y is denoted as $f_Y(\cdot)$, the two hypotheses under SIRP clutter model are expressed by

$$H_0 : f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_0) = \int_0^\infty f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_0) f_Y(y) dy, \quad (38)$$

$$H_1 : f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_1) = \int_0^\infty \lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_0) f_Y(y) dy, \quad (39)$$

where

$$\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) = \frac{f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_1)}{f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_0)} \quad (40)$$

Although H_0 in (38) is not IID, $f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | H_0)$ accomplishes property (1) (the exchangeability property) in each vector $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{iM}, x_i)$, $i=1, 2, \dots, N$, and we can define permutation statistics inside each block of $M+1$ components, corresponding to vector \mathbf{x}_i , $i=1, 2, \dots, N$, as it was done for the IID case.

The likelihood ratio of (39) and (38) is

$$\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{\int_0^\infty \lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_0) f_Y(y) dy}{\int_0^\infty f_{\mathbf{X}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_0) f_Y(y) dy} \quad (41)$$

In order to find a permutation statistic $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ equivalent to $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ of (41), we apply Proposition 2 of Section 2, taking into account that in (41) we have to consider permutations by blocks. For applying Proposition 2, we require that $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) = \phi(T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y))$ and $\phi(T; y)$ has to be an increasing function of T .

For example, in the case of Swerling II target model, we have [39]

$$f_{0i}(x|y) = \frac{x}{y^2} \exp\left(-\frac{x^2}{y^2}\right) \text{ and } f_{1i}(x|y) = \frac{x}{y^2 + S_0} \cdot \exp\left(-\frac{x^2}{2(y^2 + S_0)}\right),$$

where S_0 is the mean power of the signal return, i.e., $S_0 = E\{A^2/2\}$, being A the target signal return amplitude, and y^2 is the noise power of the corresponding Gaussian noise. Now, from (40) and (37), we have

$$\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) = \frac{\prod_{i=1}^N f_{1i}(x_i | y) \prod_{j=1}^M f_{0i}(x_{ij} | y)}{\prod_{i=1}^N f_{0i}(x_i | y) \prod_{j=1}^M f_{0i}(x_{ij} | y)} = \left(\frac{y^2}{y^2 + S_0}\right)^N \exp\left(\left(\sum_{i=1}^N x_i^2\right) \frac{S_0/2}{y^2(y^2 + S_0)}\right). \quad (42)$$

Then, $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \sum_{i=1}^N x_i^2$, and it should be compared with (24). Also, considering any block permutation in (42), we have $T_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \sum_{i=1}^N x_{ik_i}^2$, where $\mathbf{k} = (k_1, k_2, \dots, k_N)$, $k_i = 1, 2, \dots, M+1$, and $i=1, 2, \dots, N$; now, it should be compared with (25). So, the sufficient conditions of Proposition 2 in Section 2 are fulfilled, and the computable statistic $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is equivalent to $\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ of (41) for realizing the optimum permutation test. That is, the quadratic statistic is the optimum one for the permutation detector under Swerling II target models and the SIRP clutter model considered here.

For nonfluctuating target model [39], we have

$$f_{1i}(x|y) = \frac{x}{y^2} I_0\left(\frac{x\sqrt{2S}}{y^2}\right) \cdot \exp\left(-\frac{x^2 + 2S}{2y^2}\right),$$

where $I_0(\cdot)$ is the modified Bessel function of first kind and order zero [41], S is the power of the target signal return, i.e., $S = (A^2/2)$, being A the target signal return amplitude. From (40) and (37), the conditional likelihood ratio is given by

$$\lambda(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N; y) = \exp\left(-N \frac{S}{y^2}\right) \prod_{i=1}^N I_0\left(\frac{x_i \sqrt{2S}}{y^2}\right) = \exp\left(\sum_{i=1}^N \ln I_0\left(\frac{x_i \sqrt{2S}}{y^2}\right) - N \frac{S}{y^2}\right) \quad (43)$$

and there is not a computable statistic $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ that satisfies Proposition 2 of Section 2. Nevertheless, for high S -values in (43), we can find linear computable statistics: $T(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \sum_{i=1}^N x_i$ and $T_{\mathbf{k}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \sum_{i=1}^N x_{ik_i}$, where $\mathbf{k} = (k_1, k_2, \dots, k_N)$, $k_i = 1, 2, \dots, M+1$, and $i=1, 2, \dots, N$, that verify the conditions of Proposition 2 in Section 2. Consequently, the linear statistic is the asymptotically optimum one for the permutation detector under nonfluctuating target models and the SIRP clutter model considered here.

Now, the P_{fa} of the permutation test (28), under SIRP clutter, is given by

$$P_{fa} = \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | H_0\} = \int_0^\infty \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | y, H_0\} f_Y(y) dy = \frac{K}{(M+1)^N} \quad (44)$$

and the P_d of the permutation test (28), under SIRP clutter, can be expressed by

$$P_d = \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | H_1\} = \int_0^\infty \Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | y, H_1\} f_Y(y) dy, \quad (45)$$

where

$$\Pr\{T(\mathbf{X}) \geq (T_{\mathbf{k}}(\cdot))_K | y, H_1\}$$

$$= \int u(T(\mathbf{x}) - (T_{\mathbf{k}}(\cdot))_{\mathbf{k}}) f_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N | y, H_1) d\mathbf{x}_1 d\mathbf{x}_2 \dots d\mathbf{x}_N \quad (46)$$

As it has been proved above, the optimum (or sub-optimum) permutation test structure for Gaussian noise (linear or quadratic test statistic) is also optimum (or suboptimum) under the SIRP clutter considered in hypotheses (38) and (39). An alternative proof of this conclusion is as follows: for the same P_{fa} , both detectors satisfy (44) and the P_d given in (45) can be expressed by

$$P_d = \int_0^\infty (P_d(y))_{\text{Gauss}} f_Y(y) dy \quad (47)$$

where $(P_d(y))_{\text{Gauss}}$ is the detection probability (46) of the optimum (suboptimum) detector for Gaussian noise with noise power y^2 ; as $(P_d(y))_{\text{Gauss}}$ corresponds to the maximum values in the y -interval of interest, then P_d in (47) corresponds to the maximum, too. This fact was confirmed by the simulation results.

References

- [1] I.R. Savage, *Bibliography of Nonparametric Statistics*, Harvard University Press, Cambridge, MA, 1962.
- [2] S.A. Kassam, A bibliography on nonparametric detection, *IEEE Transactions on Information Theory* IT-26 (5) (Sept. 1980) 595–602.
- [3] J.D. Gibbons, S. Chakraborti, *Nonparametric Statistical Inference*, 4th edn., Marcel Dekker, Inc., New York, 2003 (1st edn. 1971).
- [4] R.H. Randles, D.A. Wolfe, *Introduction to the Theory of Nonparametric Statistics*, John Wiley & Sons, New York, 1979.
- [5] T.P. Hettmansperger, J.W. McKean, *Robust Nonparametric Statistical Methods*, 2nd edn., CRC Press (Taylor & Francis Group), Boca Raton (Florida), 2011 (1st edn. 1998).
- [6] J. Hájek, Z. Sidak, P.K. Sen, *Theory of Rank Tests*, 2nd edn., Academic Press, Inc., New York, 1999 (1st edn. 1967).
- [7] E.L. Lehmann, *Testing Statistical Hypotheses*, 2nd edn., Springer, New York, 1986 (1st edn. 1959).
- [8] P. Good, *Permutation Tests (A practical Guide to Resampling Methods for Testing Hypotheses)*, 2nd edn., Springer, New York, 2000 (1st edn. 1994).
- [9] R.A. Fisher, *The design of Experiments*, 8th edn., Oliver & Boyd, Edinburgh, 1966 (1st edn. 1935).
- [10] H. Scheffé, Statistical inference in non-parametric case, *The Annals of Mathematical Statistics* 14 (4) (1943) 305–332.
- [11] E.L. Lehmann, C. Stein, On the theory of some non-parametric hypotheses, *The Annals of Mathematical Statistics* 20 (1) (1949) 28–45.
- [12] W. Hoeffding, The large-sample power of tests based on permutations of observations, *The Annals of Mathematical Statistics* 23 (2) (1952) 169–192.
- [13] J.P. Romano, On the behavior of randomization tests without a group invariance assumption, *Journal of the American Statistical Association* 85 (411) (1990) 686–692.
- [14] W.J. Welch, Construction of permutation tests, *Journal of the American Statistical Association* 85 (411) (1990) 693–698.
- [15] J. Baglivo, M. Pagano, C. Spino, Permutation distributions via generating functions with applications to sensitivity analysis of discrete data, *Journal of the American Statistical Association* 91 (435) (1996) 1037–1046.
- [16] J.D. Gibson, J.L. Melsa, *Introduction to Nonparametric Detection with Applications*, Academic Press, Inc., New York, 1975.
- [17] S.A. Kassam, *Signal Detection in Non-Gaussian Noise*, Springer-Verlag, New York, 1988.
- [18] J.W. Carlyle, J.B. Thomas, On nonparametric signal detectors, *IEEE Transactions on Information Theory* 10 (1) (Apr. 1964) 146–152.
- [19] J.B. Thomas, Nonparametric detection, *Proceedings of the IEEE* 58 (5) (1970) 623–631.
- [20] V.G. Hansen, Detection performance of some nonparametric rank tests and an application to radar, *IEEE Transactions on Information Theory* 16 (3) (1970) 309–318.
- [21] V.G. Hansen, B.A. Olsen, Nonparametric radar extraction using a generalized sign test tests, *IEEE Transactions on Aerospace and Electronic Systems* 7 (5) (1971) 942–950.
- [22] J.L. Sanz-González, A.R. Figueiras-Vidal, A suboptimum rank test for nonparametric radar detection, *IEEE Transactions on Aerospace and Electronic Systems* 22 (6) (1986) 670–680.
- [23] J.L. Sanz-González, Nonparametric rank detectors on quantized radar video signals, *IEEE Transactions on Aerospace and Electronic Systems* 25 (6) (1990) 969–975.
- [24] J.L. Sanz-González, F. Álvarez-Vaquero, Nonparametric rank detectors under K-distributed clutter in radar applications, *IEEE Transactions on Aerospace and Electronic Systems* 41 (2) (2005) 702–710.
- [25] J. Bae, I. Song, Rank-based detection of weak random signals in a multiplicative noise model, *Signal Processing* 63 (2) (1997) 121–131.
- [26] H.G. Kim, I. Song, S.Y. Kim, J. Lee, PN code acquisition using nonparametric detectors in DS/CDMA systems, *Signal Processing* 80 (4) (2000) 731–736.
- [27] H.G. Kim, I. Song, S. Yoon, S.Y. Kim, PN code acquisition using signed-rank-based nonparametric detectors in DS/SS systems, *IEEE Transactions Vehicular Technology* 50 (4) (2001) 1151–1157.
- [28] B. Seyfe, A.R. Sharafat, Signed-rank nonparametric multiuser detection in non-Gaussian channels, *IEEE Transactions on Information Theory* 51 (4) (2005) 1478–1486.
- [29] B. Seyfe, A.R. Sharafat, Nonparametric multiuser detection in non-Gaussian channels, *IEEE Transactions on Signal Processing* 54 (1) (2006) 23–33.
- [30] F. Álvarez-Vaquero, J.L. Sanz-González, Signal processing algorithms by permutation test in radar application, *Proceedings of SPIE - The International Society for Optical Engineering* 2846 (1996) 134–140. Denver, Colorado, USA.
- [31] F. Álvarez-Vaquero, J.L. Sanz-González, A suboptimum permutation test for radar detection in log-normal clutter environments, *Proceedings of the 3rd International Conference on Information Fusion (FUSION 2000)*, Vol. 1 (Paris, France, July 10–13, 2000), MoC1-3 - MoC1-9.
- [32] J.E. González-García, J.L. Sanz-González, F. Álvarez-Vaquero, Nonparametric permutation tests versus parametric tests in radar detection under K-distributed clutter, *Proceedings of the IEEE International Radar Conference (RADAR-2005)* (Arlington, Virginia, USA, May 9–12, 2005), 250–255.
- [33] J.E. González-García, J.L. Sanz-González, F. Álvarez-Vaquero, Permutation detectors under non-homogeneous and correlated K-distributed clutter in radar applications, *Proceedings of the European Microwave Association, Special Issue on Radar Systems and Applications* 3 (1) (2007) 6–11.
- [34] J.L. Sanz-González, F. Álvarez-Vaquero, J. E. González-García, Permutation tests algorithms for nonparametric radar detection, *Proceedings of the IET International Conference on Radar Systems (RADAR 2007)*, (Edinburgh, UK, October 15–18, 2007), 328–332.
- [35] J. E. González-García, J.L. Sanz-González, F. Álvarez-Vaquero, A modified permutation test for robust radar detection under non-homogeneous and correlated clutter, *Proceedings of the International Radar Conference (RADAR' 2009)*, (Bordeaux (France), October 12–16, 2009).
- [36] L.L. Scharf, D.W. Lytle, Signal detection in Gaussian noise of unknown level: An invariance application, *IEEE Transactions on Information Theory* 17 (4) (1971) 404–411.
- [37] S. Bose, A.O. Steinhardt, Optimum array detector for a weak signal in unknown noise, *IEEE Transactions on Aerospace and Electronic Systems* 32 (3) (1996) 911–922.
- [38] S. Kraut, L.L. Scharf, R.W. Butler, The adaptive coherence estimator: A uniformly most-powerful-invariant adaptive detection statistic, *IEEE Transactions on Signal Processing* 53 (2) (2005) 427–438.
- [39] H.L. Van Trees, *Detection, Estimation and Modulation Theory (Part III)*, John Wiley & Sons, New York, 1971.
- [40] H.V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.
- [41] G.N. Watson, *A Treatise on the Theory of Bessel Functions*, 2nd edn., Cambridge University Press, England, 1966 (1st edn. 1922).
- [42] J.L. Sanz-González, F. Álvarez-Vaquero, J. E. González-García, Performance analysis of permutation and rank tests under K-distributed clutter in radar detection, *Proceedings of the International Radar Symposium (IRS-2006)*, (Kraków, Poland, May 24–26, 2006), 515–518.
- [43] J.L. Sanz-González, F. Álvarez-Vaquero, J. E. González-García, Optimum permutation and rank detectors under K-distributed clutter in radar applications, *Proceedings of the 14th European Signal Processing Conference (EUSIPCO-2006)* (Florence, Italy, September 4–8, 2006).